# Assessing the Effectiveness of Synthetic Data for Monocular Depth Estimation Models

Nicholas Haisler and Md Alimoor Reza
Drake University

## Motivation

- The goal of monocular depth estimation (MDE) is to predict pixel-wise depth from just a single input image. It has useful applications in autonomous driving, augmented reality, sensor fusion, etc [1]
- These MDE models could be cheap software alternatives to their more expensive hardware counterparts, such as LiDAR sensors.

## Problem Statement

- MDE models are effective when a large amount of ground truth data are available for training [3]. Collecting ground truth depth data for natural images is expensive.
- Studying the problem of monocular depth estimation from the perspective of training deep MDE models using synthetic data.

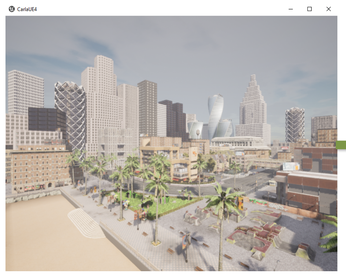
## New Synthetic Dataset

• Created an extensive collection of synthetic pairs of images and their rendered depth images using a publicly available simulator called CARLA — a testbed for autonomous driving research [4].

• Identified different attributes for scene rendering (as shown in Results section).

• Depth and RGB pairs are rendered using a python script using CARLA simulator [4].

CARLA Simulator



Depth and RGB Image Renderer

RGB

Depth

## Depth Estimation Model

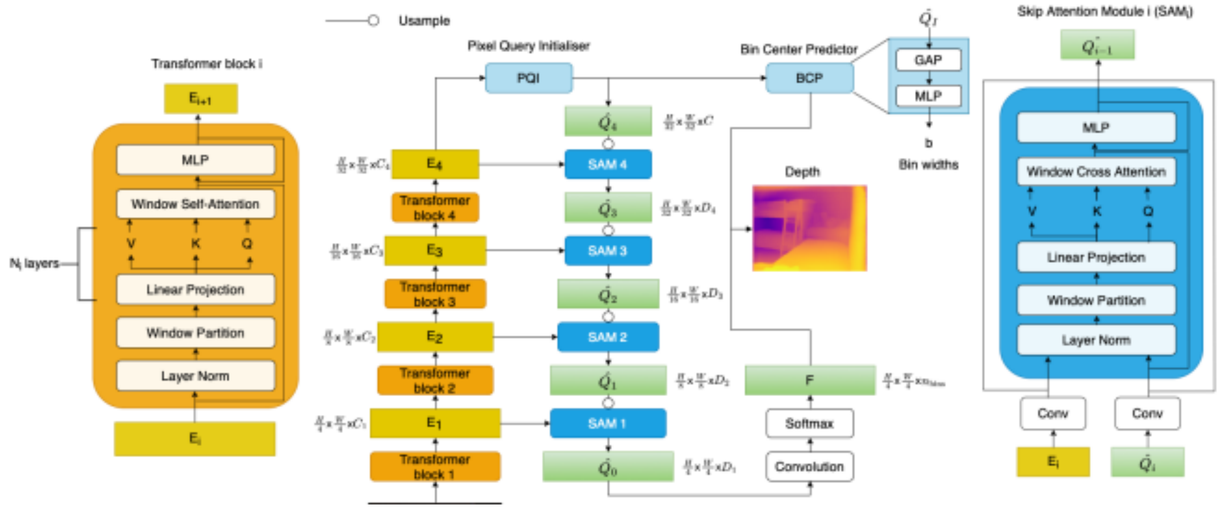



Figure: Transformer-based deep neural network architecture of PixelFormer [2] for monocular depth estimation

## Results

- Using the Carla Simulator, we generated data for training PixelFormer Model [2]
- Each row denotes a sequence of RGB and depth pairs for a particular scene. A combinations of these scenes will be used to train the depth estimation model

| Name (evan is cool) | Weather | | | | | | Human | Traffic |
|---|---|---|---|---|---|---|---|---|
| | Rain | | | | Fog | ToD | | |
| | Cloudiness | Precipitatio | Wetness | Wind Intesity | | | | |
| Images 1 | 1 | 0 | 0 | 1 | 0 | 0 | L | L |
| Images 2 | 1 | 0 | 0 | 1 | 0 | 0 | M | M |
| Images 3 | 1 | 0 | 0 | 1 | 0 | 0 | H | H |
| Images 4 | 1 | 1 | 1 | 1 | 0 | 0 | L | L |
| Images 5 | 1 | 1 | 1 | 1 | 0 | 0 | M | M |
| Images 6 | 1 | 1 | 1 | 1 | 0 | 0 | H | H |
| Images 7 | 1 | 0 | 0 | 1 | 0 | 1 | L | L |
| Images 8 | 1 | 0 | 0 | 1 | 0 | 1 | M | M |
| Images 9 | 1 | 0 | 0 | 1 | 0 | 1 | H | H |
| Images 10 | 1 | 1 | 1 | 1 | 0 | 1 | L | L |
| Images 11 | 1 | 1 | 1 | 1 | 0 | 1 | M | M |
| Images 12 | 1 | 1 | 1 | 1 | 0 | 1 | H | H |
| Images 13 | 1 | 0 | 0 | 1 | 0 | 2 | L | L |
| Images 14 | 1 | 0 | 0 | 1 | 0 | 2 | M | M |
| Images 15 | 1 | 0 | 0 | 1 | 0 | 2 | H | H |

## Future Work

- Train several depth estimation models using PixelFormer [2] on combinations of natural and synthetic data.
- Conduct experiments on the public benchmark KITTI[5][6] to assess effectiveness of the trained models using our synthetic dataset.

## References

1. D. Eigen and R. Fergus, Predicting Depth, Surface Normals and Semantic Labels with a Common Multiscale Convolutional Architecture. IEEE International Conference on computer vision (ICCV), 2015
2. A. Agarwal and C. Arora. Attention Attention Everywhere: Monocular Depth Prediction With Skip Attention. In IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023
3. R. Ranftl, A. Bochkovskiy, and V. Koltun. Vision Transformers for Dense Prediction. IEEE ICCV, 2021
4. A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez and V. Koltun. CARLA: An Open Urban Driving Simulator. Conference on Robot Learning (CoRL), 2017
5. A. Geiger, P. Lenz, and R. Urtasun. Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite, In Conference on Computer Vision and Pattern Recognition (CVPR), 2012
6. KITTI-2015 Dataset: http://www.cvlibs.net/datasets/kitti/eval_semseg.php?benchmark=semantics2015